

## Modified LeNet-5 Architecture to Classify High Variety of Tourism Object: A Case Study of Tourism Object for Education in Tinalah Village

Antonius Bima Murti Wijaya<sup>a</sup>, Desideria Cempaka Wijaya Murti<sup>b</sup>, Victoria Sundari Handoko<sup>c</sup>

<sup>a</sup> School of Computer Science, Faculty of Science and Computer, Universitas Kristen Immanuel, Indonesia

<sup>b</sup> School of Communication Science, Faculty of Social Science and Politics, Universitas Atma Jaya Yogyakarta, Indonesia

<sup>c</sup> School of Sociology, Faculty of Social Science and Politics, Universitas Atma Jaya Yogyakarta, Indonesia

Corresponding author: [bimamurti@ukrimuniversity.ac.id](mailto:bimamurti@ukrimuniversity.ac.id)

**Abstract**— This research aims to modify a CNN (Convolutional Neural Network) based on LeNet-5 to reduce overfitting in a Tinalah Tourism Village dataset object detection. Tinalah Tourism Village has many objects that can be identified for tourism education and enhanced tourist experience. While these objects, spread across the different sites of *Tinalah* do vary, some share similarities in their histogram patterns. Visually, if the size of a picture is reduced in the LeNet-5 ‘preferred size’ feature, it will inevitably lose some of its information, making pictures too similar reducing accuracy. In order to learn and classify objects, this research performs a modification on LeNet-5 architecture to provide a better performance geared toward larger input imaging. The previous state-of-the-art architecture showed an overfitting performance where the training accuracy performed too much better than the testing accuracy in our dataset. We brought in a dropout layer to reduce overfitting, increase the dense layer's size, and add a convolution layer. We then compared the modified LeNet-5 with other state-of-the-art architecture, such as LeNet-5 and AlexNet. Results showed that a modified LeNet-5 outperformed other architectures, especially in performing accuracy for testing the *Tinalah* dataset, reaching 0.913 or (91,3 %). This research discusses the dataset, the modified LeNet-5 architecture, and performance comparison between state-of-the-art CNN architecture. Our CNN architecture can be developed by involving a transfer learning mechanism to provide greater accuracy for further research.

**Keywords**— CNN; LeNet-5; image classification; digital tourism.

Manuscript received 15 Apr. 2022; revised 29 Sep. 2022; accepted 20 Dec. 2022. Date of publication 10 Sep. 2023.  
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



### I. INTRODUCTION

Tinalah Tourism Village is counted among the top 50 tourism villages in Yogyakarta province, Indonesia. The village, located in the Sungai Tinalah and Menoreh Mountain regions of Southern Java, carries a vision of becoming a nature-based education destination as part of its tourism package. The village also offers many historical sites and stories, such as the living site of Prince Diponegoro, a World War Two encryption site, as well as many geological attractions. In those sites, there are many objects and locations containing the kind of story-telling content that could be delivered to tourists to provide historical experience and knowledge.

Objects are scattered and found throughout the Tinalah village site, varying in their shape, color, and background.

There were 21 objects to be identified, including trees (coconut, cassava, bananas, *kemuning*, *kalpataru*, *cangkring*, *penjalinan*, *ketapang kencana*, and *duwet*), rocks (such as limestone, coral, and volcanic rock), traditional kitchen utensils (*tampah tumpang*, *telik*, *gejok lesung*), paintings, stalagmites, a micro-hydro machine, as well as museum items such as dioramas and historical soil. Classifying those objects is challenging due to the variety of the image class and it is a raw image with various backgrounds. Several clusters of images have high similarities inside the cluster, such as traditional kitchen utensils and trees and high diversity outside the cluster.

Furthermore, as a social impact, this research already contributed to developing a mobile apps system that performs image recognition and storyteller as one of the tourist attractions. In order to learn and classify objects via mobile application – i.e., by object photos taken by smartphone – this

paper explores a deep learning algorithm that incorporates CNN (Convolution Neural Network) as its backbone architecture.

Since LeNet 5 is claimed to have a promising performance as a CNN architecture for image recognition [1],[2]. Therefore, this research investigated the LeNet5 architecture and modified the architecture. Our Modified LeNet 5 shows promising performance and outperforms AlexNet and LeNet5 architecture. This research provides an alignment architecture for LeNet5 for Multi-Class Image classification with variety in class. This research showed how more complex deep learning architecture would also contribute to an overfitting state.

## II. MATERIALS AND METHOD

### A. CNN Development

CNN is already widespread across many sectors in object recognition due to its accuracy in performance [3], [4]. CNN also performs better than its predecessor's support vector machine-based algorithm [5]. With its ability to perform well across audio, real-time video, and image recognition alike, CNN possesses noticeable advantages over others [6]-[8]. The detection of pedestrian gait is becoming a topic of interest for CNN in real-time video processing [9], [10].

Conversely, CNN for image recognition also shows promising advances in detecting biological objects, assisting in hand gestures and sign language, plant illness, remote sensing, agriculture, and health conditions [11]–[16]. Presently, within the health sector, CNN has already provided outstanding performances in detecting

Modification and combination solutions have also been developed for CNN to help researchers reach their highest point of accuracy. Moreover, while Ensemble learning has become one of the options in providing a more stable CNN, performing voting mechanisms through multi-CNN architecture, it has resulted in higher computational efforts.[24], [25] Other modifications, with less computation involved, have been achieved by combining heuristic algorithms to adjust hyperparameters in LeNet-5, resulting in higher accuracy[26]. Involving heuristic algorithm is also used in developing CNN architecture and shows promising performance. However, it was time-consuming and needed more resources [27]. Performing transfer learning could reduce time and effort with higher performance, yet a similar model should be available first [28]. The development of the CNN architecture seems to become more realistic.

Numerous research refers to modification(s) in their backbone CNN architecture and additional layers to adapt to their specific case studies and perform a better accuracy level that lowers MSE (Mean Square Error). Deleting the convolution layer from backbone architecture, changing the activation function, modifying the pooling layer, altering the fully connected layer to SVM, and expanding the input image, also become options in a modification that have resulted in higher accuracy, whilst considering computation effort [8], [29]–[34]. Furthermore, modification for performance development in CNN architecture also happened in its activation function such as *relu* and SoftMax layer [35]. A hybrid pooling layer was also performed to increase the performance of the CNN [36].

This research performs a modification on a LeNet algorithm to provide better performance in the processing of larger input imaging and utilizes a dropout layer to reduce overfitting. We applied random rotation and random zoom techniques to increase variety in training our image algorithm in the pre-processing section. The research then investigated the data collection process in the Tourism Village to become our ground truth data and evaluated the performance of the LeNet-5 algorithm against the modified LeNet-5 architecture. By comparing our enhanced algorithm to traditional LeNet-5, we highlight improvements to the suggested architecture.

### B. CNN for Object Classification

CNN is one of the ‘most watched’ deep learning algorithms currently available to perform multi-problem task solutions such as classification, object localisation, prediction, and regression. Currently, CNN shows promising performance in organisation and image classification, with ‘classification’, for this research, defined as ‘images classified into a determined class’. Like other neural supervised network algorithms, CNN also consists of neurons that are connected with ‘weight’ at every connection. Within the architecture, CNN also employs an input layer that represents input nodes, a hidden layer function to handle complexity, and an output layer. As an additional layer, the Convolution layer is embedded in the natural neural network architecture. CNN, therefore, as can be seen, is categorized as deep learning because of the depth of the layers that make up its architecture.

The convolutional layer is a filtering layer that sums up features into an ‘n x n’ filter and converts it into different size outputs called a feature map. Like traditional neural networks, a convolution layer works involving multiplication of the input data with its ‘weight.’ Furthermore, the convolutional layer works with ‘n’ dimension of array to an ‘n’ dimension of weight, explaining why convolutional layers work for 1, 2, and 3 dimensions of array respectively. A one-dimensional array is typically used for a structured dataset, with a two-dimensional array for unstructured datasets, such as image and voice. A three-dimensional array finds its most frequent use in video processing – where time becomes an additional dimension—or when processing three-dimensional images [37]. Combinations of convolution layer’s dimensional aspects, like 2D and 3D, are also made possible to develop and enhance performance during hyperspectral image processing [38].

$$Y = \sum X_i W_i \quad (1)$$

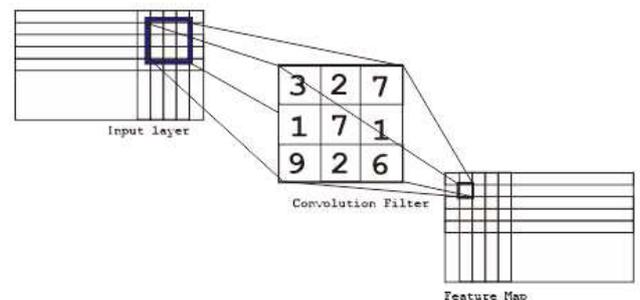


Fig. 1 Convolution Layer

Convolution works by calculating the dot operation of each filter size from input with value in the convolution filter and puts the result on the feature map (illustrated below in Figure 1). In the image classification, the input layer will consist of 3 color layers: red, green, and blue.

There are also Pooling layers where the input image will be shrunk and filtered based on the maximum, minimum, or average value from each determined padding size. Figure 2 shows a pooling layer with a 2x2 filter with three strides. The pooling layer will search for the maximum value in the filter.

The final layer is the fully connected layer. This layer will fully process the final output from convolutional and pooling layers, processing it within a dot operation, similar to traditional neural networks. In this layer, many activation functions may be employed to help determine the output from the dot matrix that has been produced, such as the Rectified Linear Unit (ReLU) and Softmax activation functions utilized in this research.

ReLU activation function as implemented in this study:

$$F(x) = \max(0, x) \quad (2)$$

Where 'x' stands for input value in the equation, anything less than 0, will be converted to 0. Alternatively, the Softmax activation function is an activation function within a neural network that helps convert a vector of numbers into a probability vector. The Softmax activation function works to calculate the probability through the exponential function:

$$F(x_i) = \frac{e^{x_i}}{\sum_{i=0}^{i=j} e^{x_i}} \quad (3)$$

Dropout is a regularization mechanism where neurons will be dropped from the neural network randomly during the training process to reduce the architecture's complexity. A complex architecture will lead to a condition known as overfitting, therefore, reducing this complexity will help the architecture to avoid overfitting.

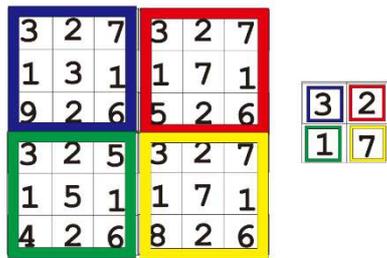


Fig. 2 Pooling layer

### C. Dataset Development and Specification

This research started by selecting suitable objects from the Tinalah Tourism Village site used for recognition. The object must have a story to tell behind it, something that can give an experience and educate while simultaneously entertaining and enhancing the tourist experience. After completing our observation and interviews with local tour guides in Tinalah Tourism Village, we determined there were 21 objects suitable set by the parameters of our research. Understanding the dataset class of each object will assist in CNN architecture development [39]. Our dataset took three different times to imbue the image with different lighting intensities. Similar images were then removed from the set because it could lead to overfitting. In order to develop different results with the

same object, we also changed the angle of shots taken and used eight different phone cameras.

TABLE I  
DATASET CLASS

No	Object Name	Picture	Story Theme
1	Coral Rock		Geological Story
2	Limestone		Geological Story
3	Volcanic Stone		Volcanology Story
4	Shadow Puppet Paint		Moral Story
5	Telik		Local Fishing tools
6	Original Sandi Site Soil		History about Sandi Site Origins
7	Tampah Tumpeng		Ceremonial Kitchen Utensil
8	Stalactite		Cave Geological Story
9	Cassava Tree		Traditional food source
10	Banana Tree		Traditional food source
11	Coconut Tree		Traditional food and hat material

No	Object Name	Picture	Story Theme
12	Penjalinan Tree		Local signature tree
13	Rice Plant		Local signature plant and food source
14	Micro hydro		Renewable Energy Source
15	Kemuning Tree		Historical Tree about Diponegoro Prince
16	Gejok Lesung		Traditional ceremonial kitchen utensil
17	Duwet Tree		Signature Tree
18	Sandi Site Diorama		Historical of Sandi Site
19	Cangkring Tree		Traditional material drugs
20	Kalpataru Tree		Tradional materials drugs
21	Ketapang Kencana Tree		Signature Tree

TABLE II  
DATASET SPECIFICATION

	Training Set	Testing Set
Number	100	20
Pixel Size	180 x 250	180 x 250
Data Size	41.1 MB	7.96 MB

This research preferred to determine pixel size in 180 x 250, larger than the LeNet-5 preferred size due to the color

and shape similarities of the trees, plants, and traditional kitchen utensils pictured. Based on Figures 3 through 6, showed that tree pictures possess a common green color histogram that is the most accurate, meaning that green, as a color, dominates each picture field. In Figures 7 to 9, red, green, and blue most closely align, showing that the kitchen utensil had color schemes mostly dominated by grey and brown tones. Therefore, maintaining the shape of an object by providing extra pixel size is important due to the similarity of color. When pictures of a similar color like plants and trees are shrunk to a smaller size, there will be less difference between pictures, making them harder to differentiate. However, it should be noted that extending pixel size has consequences when it comes to accuracy because it is here where overfitting is most likely to occur.

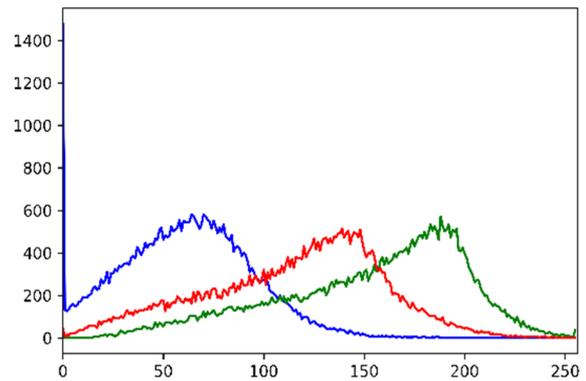


Fig. 3 Rice RGB Histogram

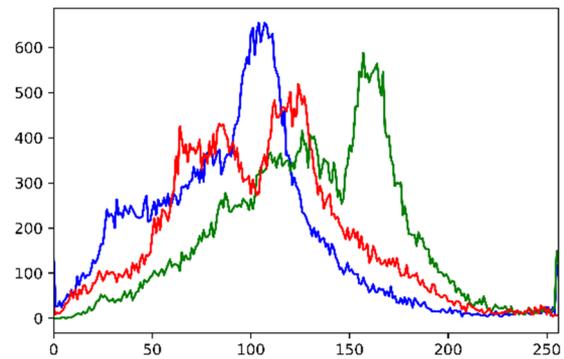


Fig. 4 Banana Tree RGB Histogram

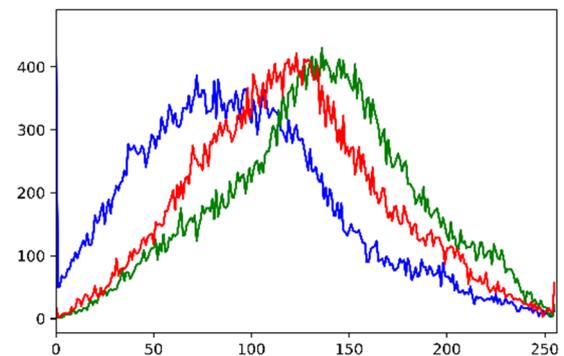


Fig. 5 Penjalinan Tree RGB Histogram

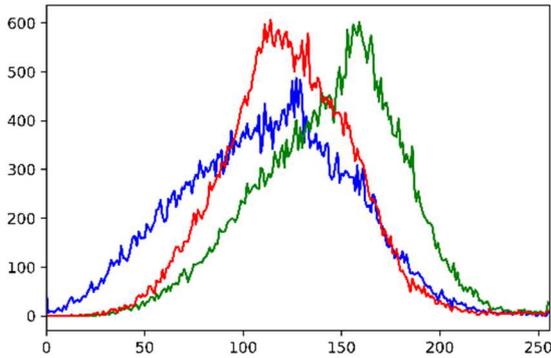


Fig. 6 Casava Tree RGB Histogram

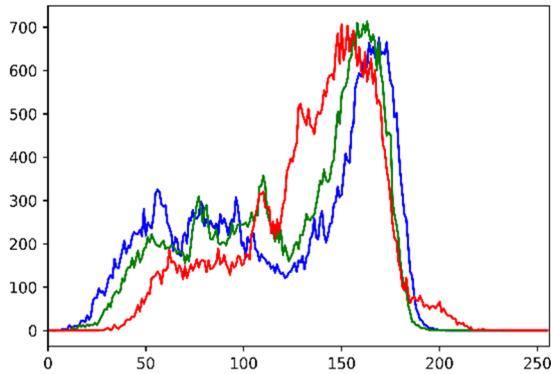


Fig. 7 Telik RGB Histogram

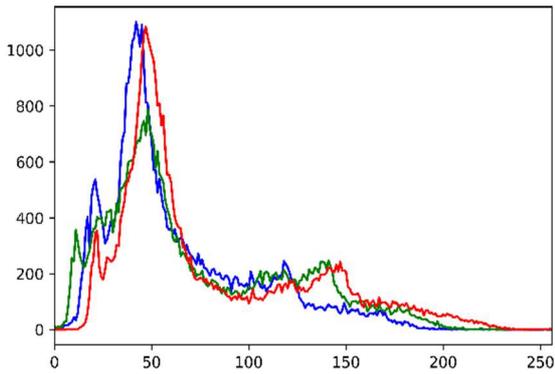


Fig. 8 Tampah Tumpeng RGB Histogram

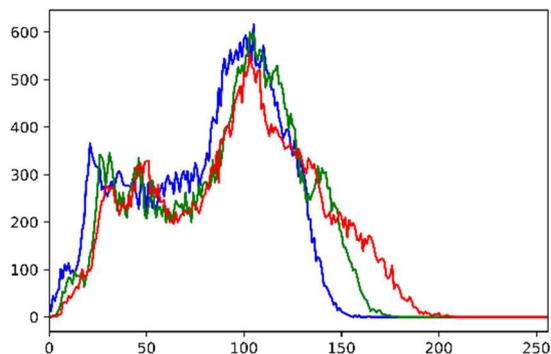


Fig. 9 Gejok Lesung RGB Histogram

#### D. The Lenet-5 Architecture Performance In Tinalah Dataset

In this research, we used LeNet-5 architecture as our backbone architecture. LeNet-5 showed an exceptional performance for image recognition [1], [2] and could be the base architecture toward developing a new and promising architecture upon modification [26], [32], [33]. Adding,

editing, and deleting become options for architecture adjustment. The performance of LeNet-5 has been shown above in Table 4. These experiments were repeated ten times, with every experiment using 150 iterations. It aims to note the best performance of the training accuracy (accuracy), to train loss (loss), to test data accuracy (value accuracy), and to test data loss (value loss). Experiment number 6 shows the best performance in value accuracy and accuracy.

TABLE III  
ARCHITECTURE OF LENET-5

Layer	Layer Type	Feature Maps	Kernel/Filter	Feature Maps Size
Input	Image	-	-	32x32
C1	2D Convolution	6	5x5	28x28
S2	Pooling	6	2x2	14x14
C3	2D Convolution	16	5x5	10x10
S4	Pooling	16	2x2	5x5
C5	2D Convolution	120	5x5	1x1
F6	Fully Connected	-	-	84
Output	Fully Connected	-	-	21

TABLE IV  
PERFORMANCE OF LENET-5

No	Best val accuracy	Best val loss	Best accuracy	Best loss	Overfitting value
1	0.8260	0.7216	0.9970	0.0094	0.1709
2	0.8618	<b>0.5523</b>	0.9980	0.0079	0.1361
3	0.8644	0.5938	0.9970	0.0101	0.1325
4	0.8618	0.6083	0.9970	0.0109	0.1351
5	0.8439	0.6840	0.9960	0.0130	0.1520
6	<b>0.8695</b>	0.5586	<b>0.9975</b>	<b>0.0101</b>	0.1279
7	0.8567	0.6359	0.9950	0.0193	0.1382
8	0.8593	0.6371	0.9965	0.0151	0.1371
9	0.8593	0.6425	0.9865	0.0468	<b>0.1272</b>
10	0.8260	0.6682	0.9870	0.0397	0.1609

As can be seen, the overfitting value or ‘gap’ between both is 0.1279447675 or 12,79447675 percent. The updates’ objective is to reduce the difference between the accuracy and value accuracy without significantly reducing the accuracy.

#### E. The Modification of Lenet-5

This research prefers to use a larger input size of 128x128 pixels. Therefore, multiple updates need to be performed to ensure maximum accuracy for the current dataset. Table 5 shows this research update in LeNet-5. Firstly, an update was made to feature map sizes in 16 in the first 2D convolution layer. Secondly, an update was also made in the second convolution layer, where the feature map size was expanded twice at 32x32. This research also added a convolution layer with a map size of 64x64. These extensions of convolution layers were developed to assist the fully connected layer process larger input imaging.

In order to handle larger image sizes, the fully connected layer needed to also to ‘bulk up’ its size. The bigger the connection utilized in the fully connected layer it was found, the better the layer’s ability to resolve higher resolution images. Consequently, the possibility of overfitting is more often likely to occur. Indeed, whilst overfitting means that the network will be an exact fit to the training data, it will, subsequently, be considered inaccurate for test data.

To resolve this problem, the current research uses a dropout layer, that randomly deletes connections that occur between two neurons so that connection sizes will be reduced but still

keep neuron numbers. However, utilizing the dropout layer may result in failure if the size of the current neuron cannot handle the accuracy. Figure 11 shows that the gap between best accuracy and val\_accuracy has a reduction rate of 0.011, where, in contrast, best accuracy fell dramatically to a level of 0.9064. Based on the ten experiments undertaken for this study, the best accuracy fell at 0,9631 found in the tenth experiment (as seen below) with a gap at 0.062, which is still less than the gap against LeNet-5 without dropout layers.

TABLE V  
ARCHITECTURE OF LENET-5

Layer	Layer Type	Val/Units	Kernel/Filter
Input	Image	-	-
C1	2D Convolution	16	3x3
S2	Avg Pooling	-	3x3
C3	2D Convolution	32	3x3
S4	Avg Pooling	-	3x3
C5	2D Convolution	64	3x3
DO1	Drop out	0.4	-
D1	Dense	120	-
DO2	Drop out	0.4	-
D2	Dense	84	-
DO3	Drop Out	0.4	-
Output	Fully Connected/Dense	21	-

In order to maintain the performance of the accuracy and minimize the gap between val\_accuracy and accuracy, this research suggests increasing the number of nodes in the fully-connected layer to 700 and 300, respectively. The dropout layer should be located in 3 places. Firstly, it is between the convolutional layer and fully connected layers to help reduce the connection within. Secondly, it is between the first and second dense layers. Thirdly, it is between the second and the output layers.

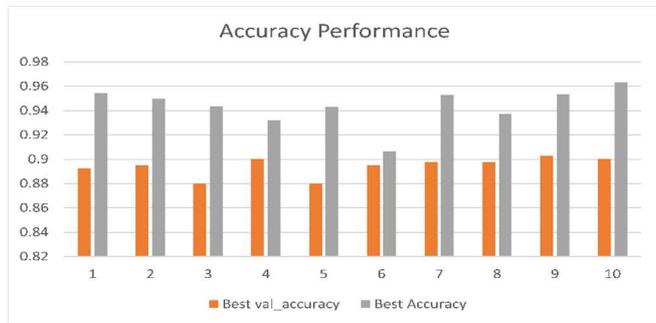


Fig. 10 Accuracy chart of the failure dropout layer

### III. RESULTS AND DISCUSSION

This LeNet-5 modification – by adapting with the expanded size and performing dropout layer – resulted in the increase in performance shown in Figures 10 and 11. Furthermore, the chart in Figure 10 demonstrates a steady accuracy value across all ten experiments, where the accuracy value did not fall too much from its original in LeNet-5. The highest accuracy reached by this algorithm is 0.993525922, performed in the first and eighth experiments. This accuracy only has 0.003984034 difference between the original LeNet algorithm. On the other hand, the performance of value accuracy increases in the proposed LeNet modification. Figure 11 shows that the best value accuracy is at 0,913043499, with the lowest at 0,897698224. This number

shows an increase in performance when compared with LeNet-5.

Figure 12 shows the difference between the value accuracy and accuracy. The best performance shown in the fifth experiment has the value reaching its lowest point at 0.078988373. It means that the architecture has less overfit, or the prediction value from the initial test data is not far from training data, with all of the overfitting value at less than that of the LeNet-5 experiment.

As noted, this research also compared the suggested architecture with other state-of-the-art architecture currently available, such as AlexNet and LeNet-5. Table 6 highlights performance between the three architectures, with the modified LeNet-5, outperforming all others in value accuracy, value loss, as well as overfitting value. Our architecture – the modified LeNet-5 – showed that it has the lowest overfitting, with a slight accuracy reduction that can be ignored. Whilst AlexNet showed promising accuracy, their gap in value accuracy is too high. Moreover, AlexNet used more connections with 4036 units in each dense layer and more convolution layer, resulting in a higher computational effort.

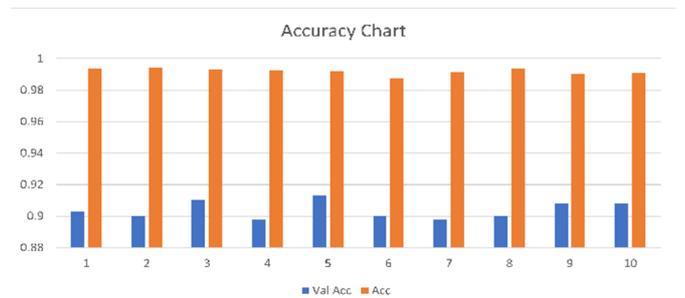


Fig. 11 Accuracy chart

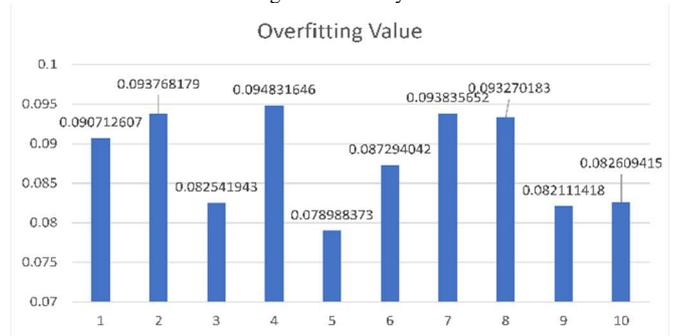


Fig. 12 Overfitting value of modified LeNet-5

TABLE VI  
PERFORMANCE COMPARISON BETWEEN ALEXNET, LENET-5 AND MODIFIED LENET-5

Algorithm	Val_Acc	Val_Loss	Acc	Loss	Overfitting Value
AlexNet	0.8132	1.1114	1	0.0000	0.1867
LeNet-5	0.8695	0.5586	0.9975	0.0101	0.1279
Modified LeNet-5	<b>0.9130</b>	<b>0.4556</b>	0.9920	0.0323	<b>0.0789</b>

### IV. CONCLUSION

LeNet-5 and AlexNet are sophisticated CNN architectures that are widely used for image recognition across the industry. In the Tinalah dataset, both show an overfitting performance where 'value accuracy that shows performance with testing

data' had an obvious difference with 'accuracy value that shows performance with training data'.

Our suggested architecture, the modified LeNet-5, outperforms both AlexNet and LeNet-5 architecture in the case of the Tinalah dataset. It shows noticeable performance differences in value accuracy, value loss, and overfitting value, highlighting important dissimilarities between value accuracy and accuracies; specifically, our architecture reaches 0.913043499 in the value accuracy 0.078988373 in the overfitting value.

When applied to Tinalah object recognition, through our chosen research, our architecture's performance needs to be increased because the desired value accuracy is more than 0.97. This performance could be reached by adding a more comprehensive data set. Furthermore, this research should also involve transfer learning to improve accuracy and reduce time execution. Transfer learning could provide a basic model for the architecture. The challenge will be how to find a proper model as the basis.

#### REFERENCES

- [1] M. Elsaadouny, J. Barowski, and I. Rolfes, "Extracting the Features of the Shallowly Buried Objects using LeNet Convolutional Network," *2020 14th European Conference on Antennas and Propagation (EuCAP)*, Mar. 2020, doi: 10.23919/eucap48036.2020.9135701.
- [2] M. Kayed, A. Anter, and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, Feb. 2020, doi: 10.1109/itce48509.2020.9047776.
- [3] X. Zhao et al., "Joint Classification of Hyperspectral and LiDAR Data Using Hierarchical Random Walk and Deep CNN Architecture," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020, doi: 10.1109/tgrs.2020.2982064.
- [4] S. Lee, Y. Sung, Y. Kim, E. Cha, "Variations of AlexNet and GoogLeNet to Improve Korean Character Recognition Performance," *Journal of Information Processing Systems*, vol. 14, no. 1, pp. 205–217, 2018. DOI: 10.3745/JIPS.04.0061.
- [5] Ş. Öztürk and U. Özkaya, "Gastrointestinal tract classification using improved LSTM based CNN," *Multimedia Tools and Applications*, vol. 79, no. 39–40, pp. 28825–28840, Aug. 2020, doi: 10.1007/s11042-020-09468-3.
- [6] Palanisamy K, Singhanian D, Yao A. Rethinking CNN models for audio classification. arXiv preprint arXiv:2007.11154. 2020.
- [7] F. Demir, M. Turkoglu, M. Aslan, and A. Sengur, "A new pyramidal concatenated CNN approach for environmental sound classification," *Applied Acoustics*, vol. 170, p. 107520, Dec. 2020, doi: 10.1016/j.apacoust.2020.107520.
- [8] Z. Bi, L. Yu, H. Gao, P. Zhou, and H. Yao, "Improved VGG model-based efficient traffic sign recognition for safe driving in 5G scenarios," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 11, pp. 3069–3080, Aug. 2020, doi: 10.1007/s13042-020-01185-5.
- [9] X. Shao, X. Nie, X. Zhao, R. Zheng, and D. Guo, "Gait Recognition based on Improved LeNet Network," *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Jun. 2020, doi: 10.1109/itnec48623.2020.9084747.
- [10] T. ul Islam, L. K. Awasthi, and U. Garg, "Gait based Age Estimation using LeNet-50 inspired Gait-Net," *International Journal of Advanced Engineering Research and Science*, vol. 8, no. 8, pp. 460–469, 2021, doi: 10.22161/ijaers.88.52.
- [11] P. P. Banik, R. Saha, and K.-D. Kim, "An Automatic Nucleus Segmentation and CNN Model based Classification Method of White Blood Cell," *Expert Systems with Applications*, vol. 149, p. 113211, Jul. 2020, doi: 10.1016/j.eswa.2020.113211.
- [12] J. Lu, L. Tan, and H. Jiang, "Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification," *Agriculture*, vol. 11, no. 8, p. 707, Jul. 2021, doi: 10.3390/agriculture11080707.
- [13] J. Qin, W. Pan, X. Xiang, Y. Tan, and G. Hou, "A biological image classification method based on improved CNN," *Ecological Informatics*, vol. 58, p. 101093, Jul. 2020, doi: 10.1016/j.ecoinf.2020.101093.
- [14] R. Vaddi and P. Manoharan, "Hyperspectral image classification using CNN with spectral and spatial features integration," *Infrared Physics & Technology*, vol. 107, p. 103296, Jun. 2020, doi: 10.1016/j.infrared.2020.103296.
- [15] A. A. Barbhuiya, R. K. Karsh, and R. Jain, "CNN based feature extraction and classification for sign language," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 3051–3069, Sep. 2020, doi: 10.1007/s11042-020-09829-y.
- [16] T.-H. Hsieh and J.-F. Kiang, "Comparison of CNN Algorithms on Hyperspectral Image Classification in Agricultural Lands," *Sensors*, vol. 20, no. 6, p. 1734, Mar. 2020, doi: 10.3390/s20061734.
- [17] M. Mateen, J. Wen, Nasrullah, S. Song, and Z. Huang, "Fundus Image Classification Using VGG-19 Architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, p. 1, Dec. 2018, doi: 10.3390/sym11010001.
- [18] G. S., V. P. Gopi, and P. Palanisamy, "A lightweight CNN for Diabetic Retinopathy classification from fundus images," *Biomedical Signal Processing and Control*, vol. 62, p. 102115, Sep. 2020, doi: 10.1016/j.bspc.2020.102115.
- [19] Y.-D. Zhang, V. V. Govindaraj, C. Tang, W. Zhu, and J. Sun, "High Performance Multiple Sclerosis Classification by Data Augmentation and AlexNet Transfer Learning Model," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 2012–2021, Dec. 2019, doi: 10.1166/jmhi.2019.2692.
- [20] H. A. Abdulkarim, M. A. Abdul Razak, R. Sudirman, and N. Ramli, "A deep learning AlexNet model for classification of red blood cells in sickle cell anemia," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 2, p. 221, Jun. 2020, doi: 10.11591/ijai.v9.i2.pp221-228.
- [21] C. Sitaula and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest X-ray image classification," *Applied Intelligence*, vol. 51, no. 5, pp. 2850–2863, Nov. 2020, doi: 10.1007/s10489-020-02055-x.
- [22] Md. R. Islam and A. Matin, "Detection of COVID 19 from CT Image by The Novel LeNet-5 CNN Architecture," *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, Dec. 2020, doi: 10.1109/iccit51783.2020.9392723.
- [23] S. A. Agnes, J. Anitha, S. I. A. Pandian, and J. D. Peter, "Classification of Mammogram Images Using Multiscale all Convolutional Neural Network (MA-CNN)," *Journal of Medical Systems*, vol. 44, no. 1, Dec. 2019, doi: 10.1007/s10916-019-1494-z.
- [24] D. Vasan, M. Alazab, S. Wassan, B. Safaei, and Q. Zheng, "Image-Based malware classification using ensemble of CNN architectures (IMCEC)," *Computers & Security*, vol. 92, p. 101748, May 2020, doi: 10.1016/j.cose.2020.101748.
- [25] H. Ismail Fawaz et al., "InceptionTime: Finding AlexNet for time series classification," *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936–1962, Sep. 2020, doi: 10.1007/s10618-020-00710-y.
- [26] Y. Zhu, G. Li, R. Wang, S. Tang, H. Su, and K. Cao, "Intelligent fault diagnosis of hydraulic piston pump combining improved LeNet-5 and PSO hyperparameter optimization," *Applied Acoustics*, vol. 183, p. 108336, Dec. 2021, doi: 10.1016/j.apacoust.2021.108336.
- [27] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3840–3854, Sep. 2020, doi: 10.1109/tycb.2020.2983860.
- [28] K. M. Hosny, M. A. Kassem, and M. M. Fouad, "Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet," *Journal of Digital Imaging*, vol. 33, no. 5, pp. 1325–1334, Jun. 2020, doi: 10.1007/s10278-020-00371-9.
- [29] C.-W. Zhang, M.-Y. Yang, H.-J. Zeng, and J.-P. Wen, "Pedestrian detection based on improved LeNet-5 convolutional neural network," *Journal of Algorithms & Computational Technology*, vol. 13, p. 174830261987360, Jan. 2019, doi: 10.1177/1748302619873601.
- [30] T. Li et al., "The image-based analysis and classification of urine sediments using a LeNet-5 neural network," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 8, no. 1, pp. 109–114, May 2019, doi: 10.1080/21681163.2019.1608307.
- [31] Q. Dai, X. Luo, and Z. Meng, "Research on Face Recognition Based on Gabor-LeNet Convolutional Neural Network Model," *Journal of Physics: Conference Series*, vol. 1650, no. 3, p. 032035, Oct. 2020, doi: 10.1088/1742-6596/1650/3/032035.

- [32] W. Cui, Q. Lu, A. M. Qureshi, W. Li, and K. Wu, "An adaptive LeNet-5 model for anomaly detection," *Information Security Journal: A Global Perspective*, vol. 30, no. 1, pp. 19–29, Aug. 2020, doi: 10.1080/19393555.2020.1797248.
- [33] W. Cui, Q. Lu, A. M. Qureshi, W. Li, and K. Wu, "An adaptive LeNet-5 model for anomaly detection," *Information Security Journal: A Global Perspective*, vol. 30, no. 1, pp. 19–29, Aug. 2020, doi: 10.1080/19393555.2020.1797248.
- [34] A. Zaibi, A. Ladgham, and A. Sakly, "A Lightweight Model for Traffic Sign Classification Based on Enhanced LeNet-5 Network," *Journal of Sensors*, vol. 2021, pp. 1–13, Apr. 2021, doi: 10.1155/2021/8870529.
- [35] R. Ahila Priyadarshini, S. Arivazhagan, M. Arun, and A. Mirmalini, "Maize leaf disease classification using deep convolutional neural networks," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8887–8895, May 2019, doi: 10.1007/s00521-019-04228-3.
- [36] T. Shanthy and R. S. Sabeenian, "Modified Alexnet architecture for classification of diabetic retinopathy images," *Computers & Electrical Engineering*, vol. 76, pp. 56–64, Jun. 2019, doi: 10.1016/j.compeleceng.2019.03.004.
- [37] M. Momeny, A. Jahanbakhshi, K. Jafarnejad, and Y.-D. Zhang, "Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach," *Postharvest Biology and Technology*, vol. 166, p. 111204, Aug. 2020, doi: 10.1016/j.postharvbio.2020.111204.
- [38] A. Rehman, M. A. Khan, T. Saba, Z. Mehmood, U. Tariq, and N. Ayesha, "Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture," *Microscopy Research and Technique*, vol. 84, no. 1, pp. 133–149, Sep. 2020, doi: 10.1002/jemt.23597.
- [39] J. Zhang, F. Wei, F. Feng, and C. Wang, "Spatial–Spectral Feature Refinement for Hyperspectral Image Classification Based on Attention-Dense 3D-2D-CNN," *Sensors*, vol. 20, no. 18, p. 5191, Sep. 2020, doi: 10.3390/s20185191.